

## 三浦文法に基づく日本語 形態素処理用文法の構築

宮崎正弘 高橋大和

新潟大学工学部情報工学科

時枝文法を発展的に継承した、関係意味論に基づく新しい日本語文法である三浦文法に基づき日本語の品詞の体系化を行った。表現に用いられる単語を、対象の種類とその捉え方に着目し、約400通り（大分類数：13）の階層化された品詞に分類して、きめ細かい品詞体系を作成した。また、規則の追加・修正が容易で拡張性に富む形態素処理用の文法記述形式を提案し、その有効性を論じた。本稿で提案した品詞体系に基づき、実際に網羅的な形態素処理用の日本語文法（ルール数：約1000）を記述した結果によれば、本文法記述形式により例外的な規則も含めて文法を簡潔に記述できるだけでなく、拡張性の点でも優れていることが分かった。

Japanese Grammar for Morphological Processing Based on Miura-Grammar

Masahiro MIYAZAKI Yamato TAKAHASHI

Dept. of Information Engineering,  
Faculty of Engineering, Niigata University

Miura-Grammar is new Japanese grammar based on the Constructive Process Theory proposed by Tokieda, and developed by Miura. This paper proposes Japanese syntactic categories system based on Miura-Grammar and formal description method of grammar rules for morphological processing. Japanese words are classified into hierarchical 400 syntactic categories. By the result of designing Japanese grammar rules for morphological processing using the proposed syntactic categories system and formal description method, it has shown that it is easy to design and improve grammar rules including non-general rules by the proposed method.

## 1. まえがき

言語表現には万人に共通する対象のあり方がそのまま表現されているわけではなく、対象のあり方が話者の認識（対象の見方、捉え方、話者の感情・意志・判断などの対象に立ち向かう話者の心的状況）を通して表現されている（言語が対象－認識－表現からなる過程的構造をもつ）ことは、国語学者・時枝誠記によって提唱された言語過程説<sup>1)～3)</sup>として知られている。時枝の言語過程説によれば、言語表現は以下のように主体的表現（辞）と客体的表現（詞）に分けられ、文は、辞が詞を重層的に包み込んだ入れ子型構造（図1参照）で表される。

- ・主体的表現：話者の主観的な感情・要求・意志・判断などを直接的に表現したものであり、日本語では助詞・助動詞（陳述を表す零記号、すなわち図1に示すように肯定判断を表すが、表現としては省略された助動詞を含む）・感動詞・接続詞・陳述副詞で表される。
- ・客体的表現：話者が対象を概念化して捉えた表現で、日本語では名詞・動詞・形容詞・副詞・連体詞・接辞で表される。主観的な感情・意志などであっても、それが話者の対象として捉えられたものであれば概念化し、客体的表現として表される。

時枝の言語過程説、およびそれに基づく日本語文法体系（時枝文法）を発展的に継承したのが三浦つとむである。三浦は、時枝の主体的表現と客体的表現の言語表現上の違いなどを継承しつつ、時枝が言語の意味を主体的意味作用（主体が対象を認識する仕方）として、話者の活動そのものに求めていたのを排し、意味は表現自体がもっている客観的な関係であるとした関係意味論<sup>3)</sup>を提唱し、それに基づく新しい日本語文法、三浦文法<sup>4)～6)</sup>を提案している。三浦文法は、細部についての分析が及んでいない部分も多々ある未完成な

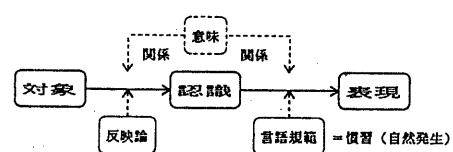
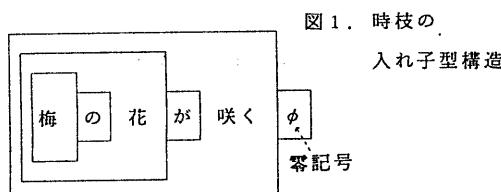
文法であるが、従来の自然言語処理の研究では見逃されていた人間の認識機構を組み込んだより高度な自然言語処理系を実現するための新しい視点を与えてくれるものと期待される。<sup>7)～10)</sup>

そこで、上記のようなより高度な自然言語処理系を実現するための第一歩として、三浦文法に基づく日本語形態素処理系を実現することを目指し、三浦文法をベースに日本語の品詞の体系化を行い、規則の追加・修正が容易で拡張性に富む形態素処理用文法を構築した。本稿では、まず三浦文法の基本的な考え方について述べ、次にそれに基づき作成した日本語の品詞体系、および品詞分類基準を示すと共に、形態素処理用の新しい文法記述形式を提案し、その有効性を論じる。

## 2. 三浦文法の言語観

時枝は、言語の本質を主体の概念作用にあるとし、言語の意味を主体の把握の仕方、すなわち対象に対する意味作用そのものとした。従って、言語表現に伴う言語規範（言語に関する社会的な約束事）とそれによる媒介の過程が無視され、認識を対象のあり方の反映とみる立場が貫かれなくなってしまい、言語による情報の伝達について、ソシュールのラングのような個人的な能力に基礎づけるところまで後退している。

これに対して、三浦は言語の意味を対象／認識／表現の関係として捉えることなど、時枝の言語過程説にいくつかの修正を加え、独自の理論的展開を図った。三浦によれば、音声や文字の種類に結び付き固定された対象と認識の客観的な関係が言語の意味である。語は使われて（表現となって）始めて意味（関係）を生じる。従って、表現が存在すれば意味は存在し、表現（文字、音声）が消滅すれば言語規範に固定されていた対象と認識の関係、すなわち意味も消滅する。対象



や認識そのものは意味ではなく、意味を形成する実体である。対象や認識が消滅しても、表現がある限り意味は存在する。意味は話者や聞き手の側にあるのではなく、言語表現そのものに客観的に存在する。三浦の言語過程説における言語モデルを図2に示す。

三浦は、時枝の提起した“主体の客体化”（1人称の代名詞は主体そのものではなく、主体が客体化されたものとみなすこと）の問題を、対象の認識の立場から発展させ、主体の観念的自己分裂と視点の移動という観点から主体を捉えた。三浦によれば、一人称の表現は見たところ、自分と話者が同一の人間であるが、これを対象として捉えているということは、対象から独立して対象に立ち向かっている別の人間（主体の観念的自己分裂によって生じた観念的話者“もう一人の自分”）が存在していると考えるのである。三浦は、このような観念的話者による視点の移動を表すものとして、観念的世界が多重化した入れ子構造の世界の中を自己分裂によって生じた観念的話者が移動する入れ子構造モデル（図3参照）を提案している。このモデルによれば、たとえば、過去の否定表現は、（過去の仮想世界／過去の現実世界／現在の現実世界）の三重の入れ子構造で表される。

### 3. 日本語の品詞体系

日本語は、膠着言語に分類される言語であり、小さな単位要素が次々と付着して表現を形成していくという特徴を持つ。これらの単位要素が結合し、表現構造を形成していく過程には一定の手順がある。言語過程説によれば、日本語の表現は客体的表現と主体的表現が入れ子になった構造として捉えることができる。

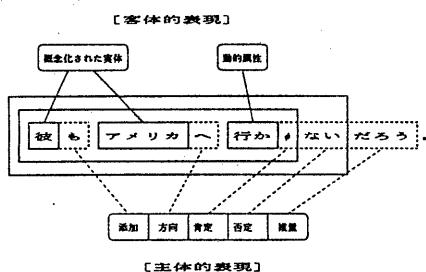


図3. 三浦の入れ子構造モデル

ここで、表現の元となる対象世界を構成する一つの事象は、実体・属性・関係の3要素から構成される。これらに対する話者の認識を言語規範を介して表現に結び付けるときに最も基本となるのは、概念化された対象（実体・属性・関係）とそれを表現する単語（詞）との対応関係、ならびに概念化された対象に対する話者自身（主体）のあり方と単語（辞）との関係である。前者に対して詞が選択され、後者においてそれに辞が付加される。このようにして概念化された対象および主体と単語との結び付きが形成されると、次にそれらの相互関係が構造化され、認識された構造と表現構造との対応づけが行われる。この過程で単語と単語が統語規則に従って構造化され、文が形成される。<sup>11)</sup>

三浦文法では他の多くの文法とは異なり、上記のような過程により形成される日本語文において、表現に用いられる単語を文構成上の機能や単語が表す内容で分類するのではなく、対象の種類とその捉え方で分類する。以下、三浦文法に基づく品詞分類の基本的考え方<sup>10)-12)</sup>について述べ、それに基づき作成した日本語の品詞体系を示す。

#### 3. 1 品詞の大分類

単語をまず以下のように客体／主体の観点から詞と辞に分ける。詞をさらに分類すると、一つの事象を表現するうえで必須である語とそうでない語の2種類がある。事象表現に必須である語は、表現の対象が実体か実体の属性かにより体言と用言に分けられる。

体言は実体・属性・関係からなる対象のうち実体を概念化したものである。実体は物理的実体と観念的実体がある。また、実体は立体的な構造を持ち、種々の側面があるため、どの側面から取り上げるかによって使用される体言も異なってくる。また、実体の構造に応じて体言間も構造的な関係を持つ。対象に立ち向かう主体（話者）も客体化して捉えた時は詞（体言）が用いられる。普通の名詞が実体のあり方を捉えたものであるのに対して、代名詞は実体と主体との特殊な関係を表現する。主体と対象の関係としては、①話者と話者の関係、②話者と聞き手の関係、③話者と話題となる事物・場所方角・人間などとの関係の3種の関係がある。実体と実体、属性と属性、実体と属性の間

には種々の関係が存在する。関係自体は、感覚的存在ではないので関係自体を概念的に対象化して名詞として用い、種々の関係は「上下（の）関係」や「親子のつながり」などのように表現することが多い。

実体の属性を概念化する用言は、動的属性を表す動詞と静的属性を表す形容詞に分けられる。属性も、これを固定的に実体化して捉えた場合は、「大きさ」や「動き」などのように名詞化する。

事象表現に必須でない語は、属性に属性を付加する副詞、および実体に属性を付加す連体詞に分類される。

話者の感情・意志・判断など対象に対する立場や対象から引き起こされる話者自身に関する認識を表す辞としては助詞や助動詞が用いられる。助詞は対象（実体）に立ち向かう話者の立場を直接表現する。「花咲く」と言えば「花」と「咲く」との間の客観的な関係を捉えたものと見ることができるが、この関係は変わらないものの、「花が咲く」「花は咲く」「花も咲く」と言えば、「花」に対する話者の立場が変化していく。このように、助詞が実体に対する話者の捉え方、すなわち、対象（もの）と主体との関係に関する主体自身の認識を表すのに対して、助動詞は対象（こと）との関係において話者自身の立場を表現するものと見ることができる。人間の認識は現実の世界だけを相手にするだけでなく、想像によって過去の世界や未来の世界、空想の世界などさまざまな世界に行き来する。このような話者の見る対象世界に対する話者の感情・意志・判断などを直接表現したものが助動詞である。

他に、話者の感情や意志などを直接表現する感動詞、話者による事象間の関係認識を表現する接続詞、および話者の主觀を強調する陳述副詞が辞に分類される。

以上の11品詞の他に、他の語（接辞承接語）に付加して別の意味や品詞性を付与する詞である接辞、文中に出現する句読点や繰返し記号などの記号類の二つを加えて、合計13の品詞に大分類した。

### 3. 2 品詞分類の細分化

単語間の文法的接続関係の検定を精密に行い、形態素処理の精度を向上させるため、品詞の大分類を細分化して品詞を約400通りに分類した。細分化のポイントを以下に示す。

#### （1）名詞

実体を同種の他の実体と共通の側面、すなわちその一般性の側面で捉えた認識を表す普通名詞、実体をその固有性の側面で捉えた認識を表す固有名詞（地名、人名、組織名、その他の固有名詞）、動的属性を固定的に実体化して捉えた認識を表す動作性名詞（サ変動詞型名詞、連用形名詞、その他の動作性転生名詞）、静的属性を固定的に実体化して捉えた認識を表す状態性名詞（静詞＜形容動詞語幹に対応・ダ型とタルト型がある＞、状態性転生名詞＜形容詞転生名詞＜例：寒さ、厚み＞＞・静詞転生名詞＜例：親切さ＞＞、連体詞型名詞＜例：大型、急性＞）、対象を具体的に取り上げることができなかつたり、取り上げる必要がない場合などに、対象を最も抽象化して捉えた認識を表す形式名詞（例：もの、こと、ため、ところ、とき、まま、際、場合）の・んく準体助詞に対応＞|ようく比況の助動詞「よう（だ）」に対応＞|そう＜伝聞の助動詞「そう（だ）」に対応＞|みたい、ふう）に細分化した。その他に特殊なものとして、具体的な数や数量など（例：1、2、…、2本）単位性の認識を表す数詞、属性に属性を付加する副詞としても用いられる副詞型名詞（時詞＜例：今日、從来＞、その他の副詞型名詞＜例：全て、みんな、多数、一部＞）を設定した。なお、静詞については格助詞（に・の・を）や肯定判断の助動詞「だ」の連体形「な」が後接するか否かを区別できるようにさらに細分化した。

#### （2）代名詞

人称代名詞（例：私、彼）、指示代名詞（例：これ、あれ）、疑問代名詞（例：だれ、どれ）の区分を導入した。

#### （3）動詞

本動詞／形式動詞の区別（3. 4の⑨参照）、活用型（五段／一段／サ変／カ変）の区別、五段動詞に対する活用行の導入により細分化した。またサ変動詞は、単独で用いる「する」「～する（例：開発する、対する）」「～する（例：論ずる）」を区別できるようにした。さらに、五段／一段動詞のうち、例外的活用をするもの（例：行く→行った、有る→有らない×、なさる→なさいます・なさい、問う→問うた、くれる→くれ）は、別品詞として区別できるようにした。

#### (4) 形容詞

ウ音便の形態（例：にくい→にくう、あさい→あそう、美しい→美しゅう）により細分化した。

#### (5) 副詞

属性をさらに具体的な面から捉えて別な語と結び付け叙述を立体化する情態副詞（例：がたがた、ピカピカ）、属性をさらに抽象的な面から捉えて別な語と結び付け叙述の程度を表す程度副詞（例：ずっと、かなり）に細分化した。さらに、格助詞（に・と・の）や肯定判断の助動詞（だ・です）が後接するか否かを区別できるよう細分化した。

#### (6) 連体詞

指示代名詞が連体詞化した指示連体詞（例：この、その）、疑問代名詞が連体詞化した疑問連体詞（例：どの、どんな）、形容詞が連体詞化した形容詞的連体詞（例：大きな）、それ以外で外延の制約を表す限定連体詞（例：ある、あらゆる）に細分化した。

#### (7) 接辞

接辞承接語との接続形態により接頭辞／接尾辞／接中辞（例：～対～）に中分類した。接頭辞については、接辞承接語の品詞により名詞接続型／動詞接続型（例：ぶち）／形容詞接続型接頭辞（例：もの／こ、ま／お）に、接尾辞については、接辞承接語＋接尾辞で構成される複合語の品詞により名詞型／動詞型（例：れる、られる、せる、させる／がる、ぶる、めく、づく、つく、じみる、過ぎる／込む、始める、終わる、続ける、くる）／形容詞型接尾辞（例：たい／らしい、がましい、(っ)ぽい／やすい、よい、にくい、づらい、がたい）に小分類した。さらに、接辞のより細かな文法的・意味的属性により以下のように細分化した。

##### 〔細分化された名詞型接頭辞とその例〕

＜普通名詞型＞県、女、核＜固有名詞型＞東、新、奥、＜動詞型＞超、反＜形容詞型＞大、新＜連体詞型＞各、全、同＜副詞型＞再、最、既＜否定型＞無、不、非、未＜前置助數詞型＞約、第＜敬意添加型＞御、ご

##### 〔細分化された名詞型接尾辞とその例〕

＜普通名詞型＞者、人、たち＜固有名詞型＞市、さん、屋、号＜動作性名詞型＞＜サ変動詞型名詞型＞化、視、＜連用形名詞型＞行き、沿い／＜その他＞発、製、＜状態性名詞型＞＜ダ型静詞型＞そう、げ、的／＜タ

ルト型静詞型＞然／＜状態性転生名詞型＞さ、み、け／＜連体詞型名詞型＞性、用、風、型、式／＜後置助數詞型＞個、番、%、＜助數詞承接型＞強、合、目／＜副詞型名詞型＞＜時詞型＞前、後、間、内、中、時、がてら／＜その他＞上／＜代名詞型＞たち、ら、自身

なお、各型の接尾辞は細分化された名詞・動詞・形容詞の品詞体系と対応付けられている（例：普通名詞型接尾辞「者」は普通名詞に対応する）。

#### (8) 接続詞

接続対象により文接続詞（例：しかし、ただし）／句接続詞（例：または、および）の区別を導入した。

#### (9) 感動詞

話者の呼びかけや感情を表す感嘆詞（例：さあ、おや、まあ）、相手の言葉に対する聞き手の応答を表す応答詞（例：はい、ええ、いいえ）に細分化した。

#### (10) 助動詞

話者の肯定判断（だ、ある、です、ます、〇／＜零判断辞＞）／否定判断（ない、ぬ、まい）／既定判断【回想・確認】（たぐたり、て）、だくだり、で）／未定判断【推量・意志】（う、よう、らしい、べし）を表す助動詞に細分化した。

#### (11) 助詞

実体のあり方の認識を表す格助詞（が、を、に、へ、と、から、より、まで、で、をば、って、して／の）、認識に対する陳述の要求を表す係助詞（は、こそ、も、さえ、すら、でも、とて、しか、しも、ぞ、して）、実体や認識に対する観念的前提の付加を表す副助詞（は／など、なんか、なんて／まで、のみ、(っ)きり、くらい、ぐらい、だけ、ばかり、ほど、とも、ずつ／や、やら、か、なり、なりと）、認識内容の確認を表す間投助詞（ね(え)、さ(あ)、よ(お)、な(あ)、の(お)／ってば、ったら、って／や、よ／だ、です、と）の他、事象間の関係づけを行う接続助詞と話者の感情を伝達する終助詞に中分類した。さらに、接続助詞は接続の型<sup>111</sup>、終助詞は伝達の方向<sup>111-133</sup>により、それぞれ、以下に示すように3通りに細分化した。

##### 〔接続助詞の細分化〕

＜同時型＞し、つつ、ながら／条件型＞ば、とも、とて、ても、でも／展開型＞と、なり、が、けれど（も）、けど（も）、に、から、ながら、ど（も）、も、して

### [終助詞の細分化]

<話者方向><<強意>>ぜ、ぞ、わ、ね(え)、さ、よ、な、とも、ってば、ったら、って、っと、い、や |  
<<驚き>>わ<相手方向><<疑問>>か、かしら、や、っけ |<<命令・勧誘>>な、ねえ、い、たら、だら |  
<<禁止>>な |<<伝聞>>と、って |<<確認>>ね(え)、さ、よ |<<婉曲>>が、けれど(も)、けど(も)<不定方向>  
<<詠嘆>>なあ、わあ、のぉ、に |<<不確定>>やら

### (12) 記号類

日本語文中に現れる記号類を、その機能に着目して  
句点相当記号（例：。？！。）、読点相当記号（例：、）、中点相当記号（例：・ <空白>）、引用符（例：「」『』‘’“”）、括弧類（例：〈〉  
《》【】〔〕〔〕）、補足記号類（例：… = -）、文頭記号（例：◎○◇▽●☆）、数式関連記号（例：、 ~ - - + - × ÷ = ≠ < > ≤ ≥ \* /）、綴返し記号（例：ゞゞゞゞ）、その他の特殊記号（例：； : @ #）に細分化した。

### 3.3 活用形の扱い

動詞、形容詞、動詞型接尾辞、形容詞型接尾辞のような活用語の活用形は、従来の学校文法における6活用形を基本とし、以下の変更を加えた。

①未然形を以下の2通りに細分化した。

- ・未然形1：推量形 [～う、～よう]
- ・未然形2：否定形 [～ぬ、～ない]

②連用形を以下の3通りに細分化した。

- ・連用形1：連用中止形 [～ます]
- ・連用修飾形
- ・連用形2：音便形 [～た、～だ]
- ・連用形3：形容詞ウ音便形 [～ござります]

③形容詞のカリ活用語尾は、以下のように扱う。

- ・かろ（未然形1）→く（形容詞語尾・連用形1）  
+ あろ（助動詞「ある」の未然形1）
- ・かっ（連用形2）→く（形容詞語尾・連用形1）  
+ あっ（助動詞「ある」の連用形2）

④タルト型形容動詞活用語尾は、以下のように扱う。

- ・と（連用形1）→と（格助詞）
- ・たる（連体形）→と（格助詞）  
+ ある（形式動詞「ある」の連体形）

### 3.4 学校文法との主な相違点

3.2～3.3で示した品詞体系と学校文法との主要な相違点は、以下の通りである。

- ①形容動詞を独立した品詞とはせず、名詞（静詞）+助動詞（肯定判断）「だ」とした。
- ②受身・使役の助動詞（れる、られる、せる、させる）は動的属性を付与する詞とし、動詞型接尾辞とした。
- ③希望の助動詞（たい）は静的属性を付与する詞とし、形容詞型接尾辞とした。
- ④伝聞の助動詞（そうだ）、比況の助動詞（ようだ）、様相の助動詞（そうだ）は助動詞とせず、それぞれ、形式名詞（そう、よう）／静詞型接尾辞（そう）+肯定判断の助動詞（だ）とした。
- ⑤準体助詞（の）、終助詞（の）は形式名詞とした。
- ⑥接続助詞（ので、のに）、終助詞（のだ）はそれぞれ、形式名詞（の）+〔格助詞（で）〕／肯定判断の助動詞（だ）の連用形1（で）]／格助詞（に）／肯定判断の助動詞（だ）とした。
- ⑦接続助詞（て、で）は既定判断の助動詞（た、だ）の連用形1とした。
- ⑧補助動詞（ある）、補助形容詞（ない）はそれぞれ、肯定判断の助動詞、否定判断の助動詞とした。
- 例：本で（は）ある／ない、  
静かで（は）ある／ない、  
重く（は）ない、重くはある、  
書いて（は）ある／ない
- ⑨既定判断の助動詞の連用形1（て、で）に後接する動詞（いる、みる、くれる、あげる、くる、もらう、やる、しまう、おく、いく、下さる、いただく、…）、形容詞連用形1／〔静詞+格助詞（に）〕に後接する動詞（する、なる）、およびサ変動詞型名詞／連用形名詞に後接する動詞（する、できる、下さる、なさる、致す、申す、申し上げる、いただく、願う、たまう、…）は、形式動詞とする〔動詞の直前に副助詞（は）などがあつてもよい〕。  
例：走って（は）いる、美しく（は）なる、  
静かに（は）なる、開発（は）する

### 4. 形態素処理用の文法記述形式

形態素処理における隣接単語間の文法的な接続検定には、通常各単語の辞書情報に前接コードと後接コードを持たせ、その二つの情報から前接コードと後接コード間の接続の可否を示すマトリックス形式の接続テーブルを用いることが多い。しかし、規則が簡潔に分かりやすく記述されておらず、例外的な接続をする単語に対しては、従来の規則との整合性を保ちながら、新しい前接・後接コードを付与したり、辞書情報を変更しなければならず、規則の追加・修正が容易でなく、規則のメンテナンス性が悪い。また、「良さそうだ」（形容詞に「そうだ」が接続する場合、通常形容詞語幹に直接「そうだ」が接続する＜例：楽しそうだ＞が、「良い・無い」は接尾辞「さ」を介して接続する）のように、2項関係だけでなく、3項関係もチェックしなければならないような例外的な接続に対応できない。

そこで、このような問題を解決するものとして、以下に述べるような形態素処理用の文法記述形式（接続ルール）を提案する。この接続ルールは基本的には、ある品詞 $p_0$ の直後（接尾辞などで直前の語の品詞などが問題となる場合には直前）に文法的に接続可能な全ての品詞 $q_1, q_2, \dots, q_m$ をリスト形式で記述し、そのリストと $p_0$ を対にして、 $((p_0)((q_1)(q_2)\dots(q_m)))$ の形で定義したもの（ルール文）の集合であり、必要に応じて3項以上の関係にも簡単に拡張できる。

文法記述においては、規則の追加・修正が容易であること、例外的な規則を記述しやすいこと、規則を簡潔に分かりやすく記述できることなどを考慮し、以下のような点を工夫した。

#### （1）接続ルールの記述量の削減

① $((s_0):((r_1)(r_2)\dots(r_n)))$ の形 [ $s_0$  : 定義文番号 (= 定義文識別子 F と通番)、 $r_i$  : 品詞] の定義文を導入し、ルール文の( $q_i$ )の位置に( $s_0$ )がある場合、 $(s_0)$ を $(r_1)(r_2)\dots(r_n)$ で置換可であることを表す。

② $((p_i)=(q_j))$ の形 [ $p_i, p_j$  : 品詞] の同格文を導入し、 $p_i$ は $q_j$ と同じ接続をすることを表す。

③文頭、文末、句境界 [句は詞（接頭辞以外の接辞・形式名詞・形式動詞は除く）、接続詞、感動詞、陳述副詞で始まる。用言・体言が後接するか否かにより4種の句境界に分類] には、文頭、文末、句境界であることを表す架空単語が存在するものとし、そ

れらに特別の品詞を設定した。それらは、ルール文、定義文などにおける規則の記述において、通常の品詞と同様に扱う。日本語では、単語間の文法的接続条件は、句内に比べて句境界では厳しくなく、句境界をまたがって接続可能な多くの品詞対がある。このような句境界をまたがって接続する品詞を句境界架空単語を介して接続するとみなすことにより、接続ルールの記述量を大幅に削減できるだけでなく、これを句境界の検出情報としても利用できる。

④4桁の16進数で表示される品詞コード（上位3桁は品詞 <1桁目は品詞の大分類を表す>、下位1桁は活用形を表す）の2～4桁目には、ワイルドカード文字（X, Y, Z <0～fの任意の値> / x, y, z: <1～fの任意の値>）が使え、これによりいくつかの品詞コードをまとめて記述できる。また、X, Y, Z, x, y, z のとなる値を $xyz/y=1, 3, 4$ のように指定できる。

#### （2）例外的な規則の記述

①品詞では表現できず、各単語に依存するような例外的規則をも前接・後接コードなどを用いて記述するため、品詞 $q_i$ と共に単語の字面（表記のゆれがある場合にはそれらの代表形である標準表記）を接続ルールに記述できるようにした。 例：( $q_i$ , “同じ”)

②形容詞のウ音便において形容詞の語幹の末尾が変化する場合（例：あさい→あそ））、単語辞書中の各単語の素性情報を用いた素性チェックにより接続判定を行う必要がある場合（例：ある形容詞語幹に接尾辞くさ、み、…が接続するか否か）など例外的処理を行うため、接続ルールからある種の手続きの起動を指示するような情報（手続き識別子 \$と手続き名）を品詞 $q_i$ と共に接続ルール中に記述できるようにした。 例：( $q_i$ , \$unbin)

③規則の適用性や優先度を示す接続確率（確率識別子 \*と確率値<0～1>、本情報を省略した場合は1とみなす）を導入し、例外的な接続をする場合にその規則の適用性が変化したり、通常、あまり適用されない規則であることなどを記述できるようにした。

例： $((p_0)((q_1)(q_2)\dots(q_i *0.5)\dots(q_m)))\dots @$

$((p_0)((q_1)(q_2)\dots(q_m)) *0.1) \dots\dots\dots @$

④：品詞 $p_0$ と品詞 $q_i$ の接続確率 = 0.5

⑤：品詞 $p_0$ と品詞 $q_i$  ( $i=1\sim m$ ) の接続確率 = 0.1

### (3) 接続ルールの拡張性

①品詞 $p_0$ に関する一般ルール( $(p_0)((q_1)(q_2)\cdots(q_m))$ )  
の他に、品詞 $p_0$ の例外ルール（単語の字面が指定されたもの、3項以上の関係を記述したものなど）がある場合、例外ルールを一般ルールの前に置いて一般ルールより高い優先順位を与え、形態素処理における接続検定において、一般ルールより先に例外ルールを先に適用し、例外ルールの適用失敗時に初めて一般ルールを適用することとした。これにより、既存の一般ルールを修正することなく例外ルールを追加することができる。また、一般ルール自身の修正も、 $(q_1)$ の変更・例外化（単語の字面指定化など）、新しい $(q_1)$ の追加などにより、他のルール文に影響を与えることなく局所的修正で済ませることができ。以上により、接続ルールの拡張が容易となった。

②接続ルールには、前接／後接コードを用いないため、これらのコードの付与、単語辞書への収録が不要となり、規則の追加、修正が容易になった。

ここで、接続ルールは人間にとて見やすく、メンテしやすく作られているが、これをそのまま形態素処理における単語接続検定に用いると、処理効率が低下する。そこで、接続ルールをこれと等価で単語接続検定が高速に行える形式の接続表 $\{(p_i, q_i, (p_i\text{の字面})(q_i\text{の字面}), (\text{接続確率}), (p_i, q_i\text{の手続き名}))\} < p_i, q_i$ はそれぞれ前接品詞・後接品詞で、ワイルドカード文字や句境界架空単語の品詞を用いず、定義文・同格文による品詞の置換済みである、（ ）は省略可、　や　の部分をそれぞれ主キー、副キーとして接続表を検索できるようになっている。3項以上の関係は別形式で記述>に変換することとし、接続ルールから接続表を自動生成するツール（接続表ジェネレータ）を作成し、形態素処理系において、接続表を用いて単語接続検定を高速に行えるようにした。

## 5. あとがき

時枝文法を発展的に継承した日本語文法である三浦文法に基づき、単語を対象の捉え方で分類することにより、日本語の品詞の体系化を行い、品詞を約400通り（大分類数：13）に分類した品詞体系を作成し

た。また、規則の追加・修正が容易で拡張性に富む形態素処理用の文法記述形式を提案し、その有効性を論じた。本稿で提案した品詞体系に基づき、本文法記述形式に従って、実際に網羅的な日本語文法（ルール数=約1000）を作成し、本文法記述形式の有効性を検証した。今後、本文法を用いた日本語形態素解析実験により本文法の検証・改良を進める予定である。

[謝辞] 三浦文法に基づく日本語品詞の体系化に関して、ご討論下さったNTT情報通信網研究所・知識処理研究部の池原悟、白井諭の両氏に深謝する。

### <参考文献>

- 1) 時枝誠記：国語学原論，岩波書店(1941)
- 2) 時枝誠記：日本文法口語篇，岩波全書(1950)
- 3) 池原悟：言語表現の意味，人工知能学会誌，vol. 6, no. 2, pp. 290-291(1991)
- 4) 三浦つとむ：認識と言語の理論，第一部～第三部，勁草書房(1967/1967/1972)
- 5) 三浦つとむ：日本語の文法，勁草書房(1975)
- 6) 三浦つとむ：日本語とはどういう言語か，講談社学術文庫(1976)
- 7) 池原, 宮崎, 白井, 林：言語における話者の認識と多段翻訳方式，情処論，vol. 28, no. 12, pp. 1269-1279(1987)
- 8) 池原, 宮崎, 白井：言語過程説から見た多段翻訳方式の意義，自然言語処理の新しい応用シンポジウム論文集，ソフトウェア科学会／電子情報通信学会，pp139-140(1992)
- 9) 宮崎正弘：言語を理解するコンピュータ・自然言語技術の展望，コンピュートロール，コロナ社，no. 37, pp. 75-81(1992)
- 10) 宮崎, 池原, 白井：言語の過程的構造と自然言語処理，8)の文献と同じ, pp. 60-69(1992)
- 11) 池原, 白井：日英機械翻訳機能試験項目の体系化，信学技報, NLC90-43, pp. 17-24(1990)
- 12) 白井, 宮崎, 池原：言語過程説から見た日本語述語の構造，8)の文献と同じ, pp. 141-142(1992)
- 13) 佐伯哲夫：語順と意味，日本語学，Vol. 2, no. 12(1983)